Theoretical Computer Science 51 (1987) 255–264
North-Holland

BOUNDED D0L LANGUAGES

Tom HEAD and Barbara LANDO

Department of Mathematical Sciences, University of Alaska, Fairbanks, AK 99775, U.S.A.

Communicated by A. Salomaa
Received November 1986

Abstract. For an alphabet A and a morphism $h: A^* \to A^*$, the set of words w such that the D0L language $L(A, h, w)$ is a BOUNDED language is shown to be B^*, where B is an effectively computable subset of A. Consequently, BOUNDEDNESS is decidable for D0L languages. The result depends on the authors' recent results on periodic D0L languages. Interpretation of the result for polynomially bounded D0L languages is also considered.

1. Introduction

A D0L system $S = (A, h, w)$ consists of a finite nonempty set A, a morphism $h: A^* \to A^*$, and a word w in A^*. The language $L(S)$ generated by S is the set $\{h^i(w) \mid i \geq 0\}$. Basic results may be found in [9, 10]. The problem under consideration is whether $L(S)$ is a BOUNDED set, where the term BOUNDED is used as in the presentation of Ginsburg [1]: X is *BOUNDED* if there are words $x_1, \ldots, x_n$ of A^* such that X is a subset of $x_1^* \ldots x_n^*$. Note that the capitalization of BOUNDED is chosen here to avoid possible confusion with a different concept having similar terminology: a D0L system (A, h, w) is *bounded by a function* $f(x)$ if, for all $i \geq 0$, the number of symbols in $h^i(w)$ is less than or equal to $f(i)$.

In determining the set $\{w$ in $A^* \mid L(A, h, w)$ is BOUNDED$\}$, a finite set B will be constructed such that $L(A, h, w)$ is BOUNDED if and only if every symbol of w is in the set B. Use will be made of results on periodic D0L systems by the authors in [5], and of results on ω-words, the infinite sequences of symbols of A, by Pansiot [8], and Harju and Linna [4].

2. Definitions and preliminary results

Let A be a finite set, and let A^* be the set of all finite strings of elements of A, including the empty string which will be denoted by 1. For each word (finite string) w of A^*, the set $\{w^i \mid i \geq 0\}$ will be denoted by w^*. By a *subword* of a word w we mean a string s such that $w = usv$ for some strings u and v. A subword s is a *suffix* of w if $w = us$, and s is a *prefix* of w if $w = sv$. In this paper, s will be called a *proper*

0304-3975/87/$3.50

suffix (or prefix) if $s \neq w$. For any word w, $|w|$ will denote the length of w and for any set A, $|A|$ will denote the cardinality of A.

A subset X of A^* is BOUNDED if there exists a finite set of words $x_1, x_2, \ldots, x_n$ in A^* such that X is a subset of $x_1^* x_2^* \ldots x_n^*$. An introduction to BOUNDED sets can be found in [1]. The following basic facts will be used:

(a) A finite product of BOUNDED sets is BOUNDED.
(b) A finite union of BOUNDED sets is BOUNDED.
(c) If X is BOUNDED and Y is a set of subwords of words in X, then Y is BOUNDED. In particular, a subset of a BOUNDED set is BOUNDED.
(d) The homomorphic image of a BOUNDED set is BOUNDED.

The proofs of (a)–(c) are found in [1, Lemma 5.1.1]. For (d), let h be a morphism $h: A^* \to A^*$ and let X be a subset of $x_1^* \ldots x_n^*$, with x_i in A^*. Then $h(X)$ is a subset of $h(x_1)^* \ldots h(x_n)^*$, and thus is BOUNDED.

A word y in A^* is *primitive* if $y = z^n$ can hold only if $y = z$ and $n = 1$. For every non-null word x there is a unique primitive word y such that $x = y^k$ for some positive k. If X is a BOUNDED set with $X \subset x_1^* \ldots x_n^*$ and, for each i, y_i is the primitive root of x_i, then X is a subset of $y_1^* \ldots y_n^*$. Thus, for any BOUNDED set X, we may assume that the factors x_i are primitive words.

Let $S = (A, h, w)$ be a D0L system with morphism $h: A^* \to A^*$ and initial word w. Let $L(S) = L(A, h, w)$ denote the associated D0L language. The following lemmas are consequences of (a)–(d) above.

Lemma 2.1. *Let (A, h, w) be a* D0L *system and let $w = uv$. $L(A, h, w)$ is* BOUNDED *if and only if $L(A, h, u)$ and $L(A, h, v)$ are* BOUNDED.

Proof. $h^i(w) = h^i(u)h^i(v)$ for each i. Thus if $L(A, h, u)$ and $L(A, h, v)$ are BOUNDED, then so is $L(A, h, w)$ since it is a subset of the product of $L(A, h, u)$ and $L(A, h, v)$. On the other hand, if $L(A, h, w)$ is BOUNDED, then so are $L(A, h, u)$ and $L(A, h, v)$ since these are sets of subwords of words of $L(A, h, w)$. □

The following immediate consequence of Lemma 2.1 is fundamental for the investigation of BOUNDED D0L languages.

Corollary 2.2. *Let $w = a_1 a_2 \ldots a_m$, where a_i is in A. Then $L(A, h, w)$ is* BOUNDED *if and only if $L(A, h, a_i)$ is bounded for each i, $i = 1, \ldots, m$.*

As a result of this corollary, we will reduce the question of whether $L(A, h, w)$ is BOUNDED to consideration of the case $L(A, h, a)$, where a is in A. It is often also convenient to 'skip ahead', that is, to consider $(A, h, h^k(w))$, $k > 0$, instead of (A, h, w), and to 'speed up', that is, to consider (A, h^i, w) with $i > 1$. The following lemma shows that these simplifications will not affect the question of BOUNDEDNESS.

Lemma 2.3. (i) *For each positive integer k, $L(A, h, w)$ is* BOUNDED *if and only if $L(A, h, h^k(w))$ is* BOUNDED.

(ii) *For each integer $i > 1$, $L(A, h, w)$ is* BOUNDED *if and only if $L(A, h^i, w)$ is* BOUNDED.

Proof. (i): Let $k > 0$ be given. Then

$$L(A, h, w) = L(A, h, h^k(w)) \cup \{w, h(w), \ldots, h^{k-1}(w)\}.$$

By (c), if $L(A, h, w)$ is BOUNDED, then so is its subset $L(A, h, h^k(w))$. By (b) and the fact that any finite set is BOUNDED, if $L(A, h, h^k(w))$ is BOUNDED, then $L(A, h, w)$ is the union of BOUNDED sets and thus is BOUNDED.

(ii): Let $i > 1$ be given, and let the D0L systems (A, h, w) and (A, h^i, w) be denoted by S and T respectively. Consider $h^j(w)$ in $L(S)$. Since $j = iq + r$, for some $q \geq 0$ and $0 \leq r < i$, $h^j(w) = h^r(h^{iq}(w))$, and therefore, $h^j(w)$ is in $h^r(L(T))$. Thus,

$$L(S) = L(T) \cup h(L(T)) \cup \cdots \cup h^{i-1}(L(T)).$$

The result follows from (c), (b), and (d) above. □

A D0L system (A, h, w) is *periodic* if there exist nonnegative integers i, p, and e, with p and e positive, such that $h^p(h^i(w)) = (h^i(w))^e$. A D0L language is periodic if it is generated by a periodic D0L system. Basic properties are given in [5, 6]. In particular, it was shown in [5] that the set

$$V = \{v \text{ in } A^* \mid v \text{ is primitive, and } h^p(v) = v^e \text{ for some } p > 0 \text{ and } e > 1\}$$

is a finite set, and its elements can be listed effectively. (See also [4] for the result in the case that $p = 1$.) The fact that the elements of V can be so listed makes the present exposition possible. The following result is an immediate consequence of Lemma 2.3 and the definition of a periodic D0L language.

Corollary 2.4. *Each periodic* D0L *language is* BOUNDED.

3. BOUNDED D0L languages

Let A be a finite set, and $h : A^* \to A^*$ a morphism. We wish to determine for which w in A^*, $L(A, h, w)$ is BOUNDED. Let $B = \{a \text{ in } A \mid L(A, h, a) \text{ is BOUNDED}\}$. By Corollary 2.2, $L(A, h, w)$ is BOUNDED if and only if w is in B^*. Thus, it suffices to determine B. We will do this by partitioning A into four subsets and determining which symbols in each subset generate BOUNDED D0L languages.

A symbol a in A is called a *finite* symbol if $L(A, h, a)$ is a finite set, and a is said to be an *infinite* symbol if $L(A, h, a)$ is infinite. Let F denote the set of finite symbols of A, and I the set of infinite symbols. Partition I into three sets:

$L = \{a$ in $I \mid$ there exists a $k > 0$ such that a occurs once in $h^k(a)$, but there is no $j > 0$ such that a occurs more than once in $h^j(a)\}$,

$M = \{a$ in $I \mid$ there exists a $k > 0$ such that a occurs more than once in $h^k(a)\}$,

$N = \{a$ in $I \mid$ there is no $j > 0$ such that a occurs in $h^j(a)\}$.

Now, $A=F\cup L\cup M\cup N$, and it can be shown that each of these subsets can be effectively determined: Specifically, if there is a $k>0$ such that a occurs in $h^k(a)$, then there is such a k with $k\leqslant|A|$; and if there is a $k>0$ such that a occurs more than once in $h^k(a)$, then there is such a k with $k\leqslant|A|^2$. Then

$$B=B\cap A=(B\cap F)\cup(B\cap L)\cup(B\cap M)\cup(B\cap N).$$

$B\cap F=F$ since, for any a in F, $L(A,h,a)$ is finite and thus BOUNDED. We proceed to the determination of $B\cap L$.

The following notation will be used. Let u be a string of A^*. Denote $h^i(u)$ by u_i, and u by u_0. We define the set $J(A,h,u)=\{uu_1\ldots u_i\mid i\geqslant 0\}$. Then, if $J(A,h,u)$ is infinite, we say that $J(A,h,u)$ is *ultimately periodic* if there are y and x in A^* such that, for each $i\geqslant 0$, $uu_1\ldots u_i$ is a prefix of yx^e for some $e\geqslant 0$. Observe that it is decidable whether $J(A,h,u)$ is ultimately periodic: Let c be a symbol not in A. Consider the D0L system $(A\cup\{c\},h',c)$ where $h'(c)=cu$ and $h'(a)=h(a)$ for all a in A. Then $J(A,h,u)$ is ultimately periodic if and only if $cuu_1u_2\ldots$ is ultimately periodic, and this can be decided via either the algorithm of Pansiot [8] or of Harju-Linna [4]. We also consider the set $\{u_i\ldots u_1u\mid i\geqslant 0\}$ and say that it is left ultimately periodic if there are x and y in A^* such that, for each $i\geqslant 0$, $u_i\ldots u_1u$ is a suffix of x^ey for some e. By defining w^{R} to be the reversal of string w in A^*, and by defining $h^{\mathrm{R}}(a)=(h(a))^{\mathrm{R}}$, we see that $\{u_i\ldots u_1u\mid i\geqslant 0\}$ is left ultimately periodic if and only if $J(A,h^{\mathrm{R}},u^{\mathrm{R}})$ is ultimately periodic.

Lemma 3.1. *Let $J(A,h,u)$ be infinite. $J(A,h,u)$ is ultimately periodic if and only if $J(A,h,u)$ is* BOUNDED.

Proof. If every word of $J(A,h,u)$ is a prefix of a word of yx^* for some y and x in A^*, then $J(A,h,u)$ consists of subwords of words of a BOUNDED set, and thus is BOUNDED.

Suppose that $J(A,h,u)\subset x_1^*\ldots x_n^*$, where each x_i is a primitive word of A^*. Let m be the minimal index such that, for all i, $uu_1u_2\ldots u_i$ is a prefix of a word of $x_1^*\ldots x_m^*$. Since m is minimal, there is a nonnegative I such that, for all $i\geqslant I$, $uu_1\ldots u_i$ is not a prefix of a word of $x_1^*\ldots x_{m-1}^*$ (where $x_1^*\ldots x_{m-1}^*$ is taken to be 1 if $m=1$). Since $J(A,h,u)$ is infinite, there is a j, $j>I$, such that $|u_{I+1}\ldots u_j|\geqslant|x_m|$. This means that any representation of $uu_1\ldots u_j$ as a prefix of a word of $x_1^*\ldots x_m^*$ is of the form $x_1^{i(1)}\ldots x_m^{i(m)}r$, where $i(m)>0$ and r is a prefix of x_m. For one such representation, denote $x_1^{i(1)}\ldots x_{m-1}^{i(m-1)}$ by y. Let $k>j$.

$$uu_1\ldots u_ju_{j+1}\ldots u_k=(yx_m^{i(m)}r)(s'(x_m^c)r'),$$

where s' is a suffix of x_m, r' is a prefix of x_m, and $c\geqslant 0$. Since x_m is primitive and $x_m^{i(m)}rs'(x_m^c)r'$ is a subword of a word of x_m^*, its representation as a subword is unique, and thus $x_m^{i(m)}rs'(x_m^c)r'$ must be a prefix of x_m^d for some $d\geqslant i(m)$. (See, for example, [6, Lemma 2.1].) Thus, $uu_1\ldots u_k$ is a prefix of yx_m^d. Therefore, every string in $J(A,h,u)$ is a prefix of a word of yx_m^*, and $J(A,h,u)$ is ultimately periodic. $\square$

Define X to be the set $\{a$ in $L \mid$ there is a $k>0$ such that $h^k(a)=uav$, and each of $J(A,(h^k)^R,u^R)$ and $J(A,h^k,v)$ is finite or ultimately periodic$\}$.

Proposition 3.2. $B \cap L = X$.

Proof. Let a be in L, with $h^k(a)=uav$. By Lemma 2.3, we may replace h^k by h, and thus assume that $h(a)=uav$. $L(A,h,a)=\{u_i\dots u_1uavv_1\dots v_i \mid i\geq 0\}$ is BOUNDED if and only if $J(A,h^R,u^R)$ and $J(A,h,v)$ are BOUNDED. By Lemma 3.1, this is the case if and only if these sets are finite or ultimately periodic. □

Example 3.3. Let $A=\{a,b,c\}$, and let h be the morphism defined on the elements of A by

$$a\to ab,\qquad b\to bc,\qquad c\to c.$$

Both a and b are in L. Since $b\to bc$ and $J(A,h,c)=\{c^i \mid i>0\}$ is ultimately periodic, $L(A,h,b)$ is BOUNDED (Proposition 3.2). For $a\to ab$, we consider $J(A,h,b)=\{bc^0bc^1\dots bc^i \mid i\geq 0\}$. Since $J(A,h,b)$ is clearly not ultimately periodic, $L(A,h,a)$ is not BOUNDED.

Next we seek a characterization of $B\cap M$. Elementary facts about conjugate words will be used. Recall that two words u and v in A^* are *conjugate* if there is a word z in A^* such that $uz=zv$. Let $\sim$ denote 'is a conjugate of'. The following lemma depends on the observation that if $u\sim v$, then $h(u)\sim h(v)$.

Lemma 3.4. *Let u be in A^*, and $h(u)\sim u^f$ for some $f>0$. Then there is a conjugate v of u and integers $p>0$ and $e\geq f$ such that $h^p(v)=v^e$.*

Proof. Denote u by v_0, and let $h(v_0)=v_1^f$, where $v_1\sim v_0$ and $f>0$. Since $v_1\sim v_0$, $h(v_1)=v_2^f$, where $v_2\sim v_1\sim v_0$. Continuing in this manner we obtain a sequence v_0, v_1, $v_2,\dots$ of conjugates of v_0. Since the number of conjugates is finite, there are integers i and j, with $i<j$, such that $v_i=v_j$. Then $h^{j-i}(v_i)=(v_j)^{f^{j-i}}=(v_i)^{f^{j-i}}$. Thus we obtain the result with $p=j-i>0$ and $e=f^{j-i}\geq f$. □

As noted in Section 2, the set $V=\{v$ in $A^* \mid v$ is primitive and $h^p(v)=v^e$ for some $p>0$ and $e>1\}$ is finite and effectively computable. Let $Y=\{a$ in $I \mid a$ occurs in some v in $V\}$. Note that Y is also an effectively computable set.

Proposition 3.5. $B\cap M = Y$.

Proof. ($Y\subset B\cap M$): Let a be in Y. Then there is a v in V such that a occurs in v. Let p be the period of v so that $h^p(v)=v^e$, where $e>1$. Since a occurs in v, $h^{ip}(a)$ is a subword of $h^{ip}(v)=v^{e^i}$, for all i. Thus, $L(A,h^p,a)$ is BOUNDED, and hence $L(A,h,a)$ is BOUNDED. Since a is an infinite symbol, there is a $q>0$ such that $h^{qp}(a)=sv^2r$, with a occurring at least twice. Thus, a is in M.

($B\cap M\subset Y$): Let a be in $B\cap M$, with $L(A,h,a)\subset x_1^*\dots x_n^*$, where each x_i is a primitive word of A^*. Speed up so that, for all $i>0$, a occurs in $h^i(a)$. Observe that there is an upper bound on the length of subwords of $h^i(a)$ in which a does not occur: Suppose not. Then there is a $k>0$ such that $h^k(a)=w_1uw_2$ where $|u|>|x_j|$ for all j, and there is exactly one occurrence of the symbol a in u and that occurrence is either as the initial symbol of u or as the terminal symbol of u. Since a is in M, there is a q such that the number of occurrences of a in $h^q(a)$ exceeds n.

$$h^kh^q(a)=h^k(z_0az_1a\dots z_naz_{n+1})=h^k(z_0)(w_1uw_2)h^k(z_1)(w_1uw_2)\dots h^k(z_{n+1}).$$

Since $h^k(h^q(a))$ is in $x_1^*\dots x_n^*$ and a occurs only once in u, each occurrence of u must overlap two different x_j's. Thus, there must be at least $n+1$ different x_j's: a contradiction. Therefore, there is a bound on the length of subwords in which a does not occur.

We now inspect the factors x_i. Consider a specific sequence of representations of the $h^k(a)$ as words of $x_1^*\dots x_n^*$:

$$\begin{aligned} h^0(a)&=x_1^{i(1,0)}x_2^{i(2,0)}\dots x_n^{i(n,0)},\\ h^1(a)&=x_1^{i(1,1)}x_2^{i(2,1)}\dots x_n^{i(n,1)},\\ &\vdots\\ h^k(a)&=x_1^{i(1,k)}x_2^{i(2,k)}\dots x_n^{i(n,k)},\dots, \end{aligned}\tag{$*$}$$

where the $i(j,k)\geq 0$. We will call x_j a bounded factor if $\{i(j,k)\mid k=0,\dots\}$ is bounded, and we will call x_j an unbounded factor otherwise. Since a is an infinite symbol, there must be at least one unbounded factor. By the previous paragraph, every unbounded factor must contain an occurrence of a. Thus, for every unbounded factor u, $h^k(u)$ grows in length as k gets large. Speed up so that, for every unbounded factor u and for all j, $j=1,\dots,n$, $|h(u)|>2|x_j|$ (and assume appropriate re-indexing in the specific representation ($*$) above).

We next show that, for any unbounded factor u, there is an unbounded factor u' and an integer $e>1$ such that $h(u)\sim(u')^e$. Since u is unbounded, for any $m>0$, u^m and $h(u^m)$ are subwords of words of $L(A,h,a)$. In particular, let $b=\max\{2,$ bound of any bounded factor$\}$, and consider u^{nb}. $h(u^{nb})$ is a subword of $h^k(a)$ for some $k\geq 0$. Thus, in representation ($*$), $h(u^{nb})$ is a subword of $x_1^{i(1,k)}\dots x_n^{i(n,k)}$. $h(u^{nb})=(h(u)^b)^n$, and at least one of the n factors of the form $h(u)^b$ is a subword of a single factor $x_j^{i(j,k)}$. Since $|h(u)^b|>|x_j^{2b}|$, x_j must be an unbounded factor, say u'. Since $|h(u)|>2|u'|$, $h(u)=s(u')^cr$, where s is a proper suffix of u', r is a proper prefix of u', and $c>0$. But, since $h(u)^2$ is also a subword, $h(u)^2=s'(u')^dr'$, where s' is a proper suffix of u', r' is a proper prefix of u', and $d>1$. Thus,

$$h(u)^2=s'(u')^dr'=s(u')^crs(u')^cr.$$

Since u' is primitive, factorizations of this type are unique, and thus rs must be u' or 1. Therefore, $h(u)=s(u')^cr\sim(u')^e$, where $e>1$.

It now follows that, for some unbounded u, $h^q(u) \sim u^f$ for some $q > 0$ and $f > 1$. Let u_0 be an unbounded factor. By the above reasoning, we may obtain a sequence $u_0, u_1, \ldots,$ of unbounded factors with $h(u_i) \sim u_{i+1}^{e(i)}$. But, there are only a finite number of factors, so, for some i and $q > 0$, $u_{i+q} = u_i$ and hence, $h^q(u_i) \sim u_i^f$, where $f = e(i) \ldots e(i+q-1)$. Then, by Lemma 3.4, there is a conjugate v of u_i such that $h^{pq}(v) = v^e$ for some $p > 0$ and $e \geq f$. Since v is a conjugate of u_i, v is primitive and contains an occurrence of a. Thus, a is in Y. $\square$

Example 3.6. Let $A = \{a, b, c, d\}$ and let h be defined on the symbols of A by

$$a \to ab, \qquad b \to ca, \qquad c \to bc, \qquad d \to dad.$$

Then $M = \{a, b, c, d\}$. Since $v = abc$ is primitive and $h(v) = v^2$, v is in V. Let u be any other word of V, with $h^p(u) = u^e$. By [4, Theorem 1], applied to h^p, u must be a conjugate of abc or consist of infinite symbols other than a, b, and c. Thus, d cannot appear in any word of V, so $Y = \{a, b, c\} = B$. Furthermore, since a, b, and c are all symbols of the same word abc of V, $L(A, h, a)$, $L(A, h, b)$, and $L(A, h, c)$ are all subsets of the set $b^*c^*(bc)^*(abc)^*(ab)^*a^*$.

Finally, we obtain a characterization of $B \cap N$. Let Z_0 be the empty set. For each positive integer $n+1$ let

$$Z_{n+1} = \{a \text{ in } N \mid h(a) \text{ is in } (F \cup X \cup Y \cup Z_n)^*\}.$$

Then $Z_0 \subset Z_1 \subset Z_2 \subset \cdots$ is a nest of subsets of the set N. Consequently, there is a least positive integer $k \leq |N|$ for which $Z_{k+1} = Z_k$. Let $Z = Z_k$.

Proposition 3.7. $B \cap N = Z$.

Proof. ($Z \subset B \cap N$ by finite induction): We have $Z_0 \subset B \cap N$ and we suppose that for a nonnegative integer i we have $Z_i \subset B \cap N$. If $Z_{i+1} = Z_i$, we have $Z_{i+1} \subset B \cap N$. Suppose that $Z_{i+1} \neq Z_i$ and that a is any symbol in $Z_{i+1} - Z_i$. Let $h(a) = b_1 b_2 \ldots b_n$. Then

$$L(A, h, a) \subset \{a\} \cup L(A, h, b_1) L(A, h, b_2) \ldots L(A, h, b_n)$$

is BOUNDED, i.e., a is in B. Consequently, $Z_{i+1} \subset B \cap N$, and, by induction, $Z \subset B \cap N$.

($Z = B \cap N$ by contradiction): Let $p = |(B \cap N) - Z|$. Suppose that p is positive. Let a be *any* symbol in $(B \cap N) - Z$, and let b be *any* symbol occurring in $h(a)$. Then b is in B. By Proposition 3.2 (respectively Proposition 3.5) if b is in L (respectively M), then b is in X (respectively Y). Since a is in $(B \cap N) - Z$, $h(a)$ must contain a symbol c that is not in $F \cup X \cup Y \cup Z$. But then c cannot be in $F \cup L \cup M \cup Z$ either, and so c, like a, is in $(B \cap N) - Z$. Since a is in N, $c \neq a$. A contradiction now follows in p steps: Let a_0 be any symbol in $(B \cap N) - Z$. For each positive integer i, $h(a_{i-1})$ contains a symbol a_i in $(B \cap N) - Z$ that is different from the i distinct symbols $a_0, a_1, \ldots, a_{i-1}$. $\square$

Example 3.8. Let $A = \{a, b, c, d\}$ and let h be defined on symbols of A by

$$a \to bc, \qquad b \to bc, \qquad c \to d^2, \qquad d \to d^2.$$

Then $X = \{b\}$, $Y = \{d\}$, and $Z = \{a, c\}$. In the construction of Z, first c is in Z_1 since $h(c) = d^2$ is in Y^*, and then a is in Z_2 since $h(a) = bc$ is in $(X \cup Z_1)^*$.

The results obtained are summarized in the statement of the following theorem.

Theorem 3.9. *Let A be a finite set, $h: A^* \to A^*$ be a morphism, and w be a word of A^*. $L(A, h, w)$ is* BOUNDED *if and only if w is in B^*, where B is the union of four disjoint subsets of A defined as follows*:
- *F: the set of finite symbols*;
- *X: the set of infinite symbols a such that*
  (i) *there is one occurrence of a in $h^k(a)$ for some $k > 0$, and no more than one a in any $h^j(a)$*,
  (ii) *for $h^k(a) = uav$, the sets*

$$\{(h^i(u) \dots h(u)u)^R \mid i \geq 0\} \quad \text{and} \quad \{vh(v) \dots h^i(v) \mid i \geq 0\}$$

are finite or ultimately periodic;
- *Y: the set of infinite symbols a such that a occurs in a word of $V = \{v$ in $A^* \mid v$ is primitive and $h^p(v) = v^e$ for some $p > 0$ and $e > 1\}$*;
- *Z: the set constructed as follows*:
  (i) *let N be the set of infinite symbols a such that a does not occur in $h^k(a)$ for any $k > 0$, and let Z_0 be the empty set*,
  (ii) *for each positive integer $n+1$, let $Z_{n+1} = \{a$ in $N \mid h(a)$ is in $(F \cup X \cup Y \cup Z_n)^*\}$*,
  (iii) *let j be the least positive integer such that $Z_{j+1} = Z_j$, and let $Z = Z_j$.*

Since all of the sets F, X, Y, and Z of the theorem are effectively computable, we have the following corollary.

Corollary 3.10. *For any* D0L *system (A, h, w), it is decidable whether $L(A, h, w)$ is* BOUNDED.

4. Interpretation for polynomially bounded D0L systems

A D0L system (A, h, w) is said to be *polynomially bounded* if there is a polynomial $p(x)$ such that, for each $n \geq 0$, $|h^n(w)| \leq p(n)$. The system is *linearly bounded* if $p(x)$ is a linear polynomial. The structure of polynomially bounded D0L systems was studied by Ehrenfeucht and Rozenberg in [2], briefly reported on in [3], and discussed by Head and Wilkinson in [7]. It follows from these works (explicitly from [7, Proposition 5.1]) that (A, h, w) is linearly bounded if and only if there is a nonnegative integer K such that, for all $i \geq 0$, $h^i(w)$ has no more than K occurrences of infinite symbols. The final result relates BOUNDEDNESS to linearly bounded systems, with a proof that reviews concepts introduced in the previous sections.

Theorem 4.1. *Let $S = (A, h, w)$ be a polynomially bounded* D0L *system. $L(S)$ is* BOUNDED *if and only if S is linearly bounded.*

Proof. It will suffice to consider the case $S = (A, h, a)$ where a is symbol of A and S is polynomially bounded. The notation of Section 3 will be used, with $A = F \cup L \cup M \cup N$, and $B = F \cup X \cup Y \cup Z$. First assume that $L(S)$ is BOUNDED, with $L(S) \subset x_1^* \ldots x_n^*$. Suppose that S is not linearly bounded. Then there is an infinite symbol b_0 which occurs at least n times in some word w_0 of $L(S)$. If b_0 does not occur in $h^k(b_0)$ for any $k > 0$, then $h(b_0)$ must contain a different infinite symbol b_1, which must occur at least n times in $w_1 = h(w_0)$. Since there are only a finite number of infinite symbols, we must find some infinite symbol b such that b occurs in $h^k(b)$ for some $k > 0$, and b occurs at least n times in some word w of $L(S)$. Since b cannot be exponential, b must be in the set L. Let k be sufficiently large so that $h^k(b) = ubv$, and $|ubv| > 2|x_i|$, for all i, $i = 1, \ldots, n$. Then, $h^k(w)$ has at least n occurrences of ubv, and hence must involve at least $n+1$ different factors x_i. But this is a contradiction. Thus, S is linearly bounded.

On the other hand, assume that $S = (A, h, a)$ is linearly bounded. We note that no symbols of M are polynomially bounded, and thus, a is in F or L or N. If a is in L, then $h^k(a) = uav$, and [7, Lemma 5.4] states that u and v must consist of finite symbols. In this case the sets $J(A, (h^k)^R, u^R)$ and $J(A, h^k, v)$ are finite or ultimately periodic, and thus a is in X.

It remains to be shown that if a is in N, then a is in Z. First note that if b is a symbol of $h^j(a)$ for any $j > 0$, then (A, h, b) is linearly bounded and so b is in $F \cup X \cup N$. Now suppose that a is in $N - Z$. Then there must be a symbol a_1 occurring in $h(a)$ which is in $N - Z$ and which is different from a. In fact, for each positive integer i, there must be a symbol a_i occurring in $h(a_{i-1})$ which is different from any of $a, a_1, \ldots, a_{i-1}$. This contradicts the finiteness of N, and thus a must be in Z. Thus, it has been shown that if (A, h, a) is linearly bounded, then a is in $F \cup X \cup Z$, and hence $L(A, h, a)$ is BOUNDED. $\square$

References

[1] S. Ginsburg, *The Mathematical Theory of Context-Free Languages* (McGraw-Hill, New York, 1966).

[2] A. Ehrenfeucht and G. Rozenberg, On the structure of polynomially bounded D0L systems, Tech. Rept. No. CU-CS-023-73, Dept. of Computer Science, Univ. of Colorado, Boulder, CO, 1973.

[3] A. Ehrenfeucht and G. Rozenberg, D0L systems with rank, in: G. Rozenberg and A. Salomaa, eds., *L Systems* (Springer, New York, 1974).

[4] T. Harju and M. Linna, On the periodicity of morphisms on free monoids, *Theoret. Inform. Appl.* **20** (1986) 47–54.

[5] T. Head and B. Lando, Periodic D0L languages, *Theoret. Comput. Sci.* **46** (1986) 83–89.

[6] T. Head and B. Lando, Regularity of sets of initial strings of periodic D0L systems, *Theoret. Comput. Sci.* **48** (1986) 101-108.
[7] T. Head and J. Wilkinson, Code properties and homomorphisms of D0L systems, *Theoret. Comput. Sci.* **36** (1985) 295-312.
[8] J.J. Pansiot, Decidability of periodicity for infinite words, *Theoret. Inform. Appl.* **20** (1986) 43-46.
[9] G. Rozenberg and A. Salomaa, *The Mathematical Theory of L Systems* (Academic Press, New York, 1980).
[10] A. Salomaa, *Jewels of Formal Language Theory* (Computer Science Press, Rockville, MD, 1981).